

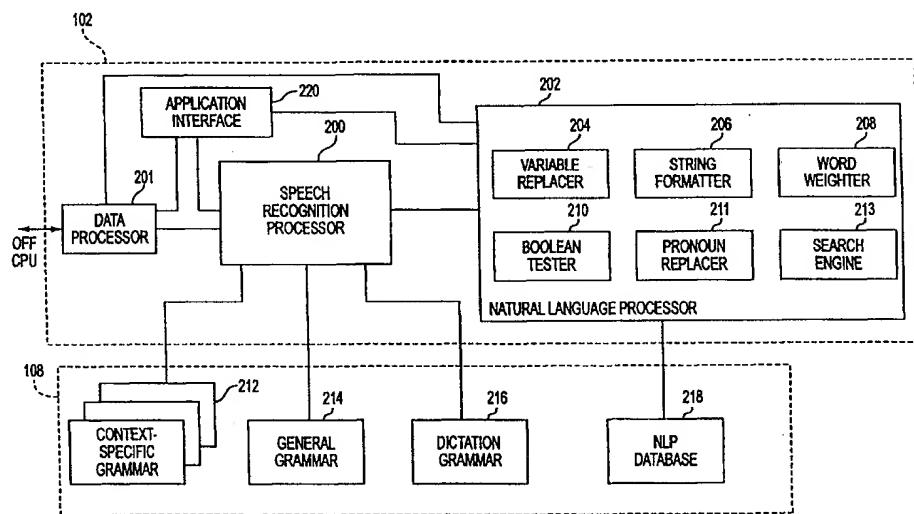


INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :	A1	(11) International Publication Number: WO 00/14728
G10L 15/26, H04L 29/06, G06F 17/27, 3/16		(43) International Publication Date: 16 March 2000 (16.03.00)

(21) International Application Number: PCT/US99/20447	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 8 September 1999 (08.09.99)	
(30) Priority Data:	
09/150,459 09/166,198	9 September 1998 (09.09.98) US 5 October 1998 (05.10.98) US
(71) Applicant (for all designated States except US): ONE VOICE TECHNOLOGIES, INC. [US/US]; Suite 240, 6333 Greenwich Drive, San Diego, CA 92122 (US).	
(72) Inventor; and	
(75) Inventor/Applicant (for US only): WEBER, Dean, C. [US/US]; 11422 Cypress Terrace Place, San Diego, CA 92131 (US).	
(74) Agents: PAULSON, Mark, G. et al.; Pillsbury Madison & Sutro LLP, 1100 New York Avenue, N.W., Washington, DC 20005 (US).	

(54) Title: NETWORK INTERACTIVE USER INTERFACE USING SPEECH RECOGNITION AND NATURAL LANGUAGE PROCESSING



(57) Abstract

A system and method for interacting with networked objects, via a computer using utterances, speech processing and natural language processing. A Data Definition File relates networked objects and a speech processor. The Data Definition File encompasses a memory structure relating the networked objects, including grammar files and a natural language processor. The speech processor searches a first grammar file for a matching phrase for the utterance, and for searching a second grammar file for the matching phrase if the matching phrase is not found in the first grammar file. The system also includes a natural language processor for searching a data base for a matching entry for the matching phrase; and an application interface for performing an action associated with the matching entry if the matching entry is found in the database. The system utilizes context-specific grammars, thereby enhancing speech recognition and natural language processing efficiency. Additionally, the system adaptively and interactively "learns" words and phrases, and their associated meanings.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**NETWORK INTERACTIVE USER INTERFACE USING SPEECH
RECOGNITION AND NATURAL LANGUAGE PROCESSING**

BACKGROUND OF THE INVENTION

5 I. Field of the Invention

The present invention relates to speech recognition for a network computer user interface. More specifically, the present invention relates to a novel method and system for user interaction with a computer using speech recognition and natural language processing. This application is a continuation-in-part of U.S. Patent Application Serial No. 10 09/150,459, entitled "Interactive User Interface Using Speech Recognition and Natural Language Processing," filed September 10, 1998.

II. Description of the Related Art

As computers have become more prevalent it has become clear that many people have great difficulty understanding and communicating with computers. A user must often learn archaic commands and non-intuitive procedures in order to operate the computer. For example, most personal computers use windows-based operating systems which are largely menu-driven. This requires that the user learn what menu commands or sequence of commands produce the desired results.

Furthermore, traditional interaction with a computer is often slowed by manual input devices such as keyboards or mice. Many computer users are not fast typists. As a result, much time is spent communicating commands and words to the computer through these manual input devices. It is becoming clear that an easier, faster and more intuitive method of communicating with computers and networked objects, such as web-sites, is needed.

One proposed method of computer interaction is speech recognition. Speech recognition involves software and hardware that act together to audibly detect human speech and translate the detected speech into a string of words. As is known in the art, speech recognition works by breaking down sounds the hardware detects into smaller non-divisible sounds called phonemes. Phonemes are distinct units of sound. For example, the word "those" is made up of three phonemes; the first is the "th" sound, the second is the "o" sound, and the third is the "s" sound. The speech recognition software attempts to match the detected phonemes with known words from a stored dictionary. An example of a speech recognition system is given in U.S. Patent No. 4,783,803, entitled "SPEECH RECOGNITION APPARATUS AND METHOD", issued November 8, 1998, assigned to

Dragon Systems, Inc., and incorporated herein by reference. Presently, there are many commercially available speech recognition software packages available from such companies as Dragon Systems, Inc. and International Business Machines, Inc.

One limitation of these speech recognition software packages or systems is that they typically only perform command and control or dictation functions. Thus, the user is still required to learn a vocabulary of commands in order to operate the computer.

A proposed enhancement to these speech recognition systems is to process the detected words using a natural language processing system. Natural language processing generally involves determining a conceptual "meaning" (e.g., what meaning the speaker intended to convey) of the detected words by analyzing their grammatical relationship and relative context. For example, U.S. Patent No. 4,887,212, entitled "PARSER FOR NATURAL LANGUAGE TEXT", issued December 12, 1989, assigned to International Business Machines Corporation and incorporated by reference herein teaches a method of parsing an input stream of words by using word isolation, morphological analysis, dictionary look-up and grammar analysis.

Natural language processing used in concert with speech recognition provides a powerful tool for operating a computer using spoken words rather than manual input such as a keyboard or mouse. However, one drawback of a conventional natural language processing system is that it may fail to determine the correct "meaning" of the words detected by the speech recognition system. In such a case, the user is typically required to recompose or restate the phrase, with the hope that the natural language processing system will determine the correct "meaning" on subsequent attempts. Clearly, this may lead to substantial delays as the user is required to restate the entire sentence or command. Another drawback of conventional systems is that the processing time required for the speech recognition can be prohibitively long. This is primarily due to the finite speed of the processing resources as compared with the large amount of information to be processed. For example, in many conventional speech recognition programs, the time required to recognize the utterance is long due to the size of the dictionary file being searched.

An additional drawback of conventional speech recognition and natural language processing systems is that they are not interactive, and thus are unable to cope with new situations. When a computer system encounters unknown or new networked objects, new relationships between the computer and the objects are formed. Conventional speech recognition and natural language processing systems are unable to cope with the situations

that result from the new relationships posed by previously unknown networked objects. As a result, a conversational-style interaction with the computer is not possible. The user is required to communicate complete concepts to the computer. The user is not able to speak in sentence fragments because the meaning of these sentence fragments (which is dependent 5 on the meaning of previous utterances) will be lost.

What is needed is an interactive user interface for a computer which utilizes speech recognition and natural language processing which avoids the drawbacks mentioned above.

SUMMARY OF THE INVENTION

10 The present invention is a novel and improved system and method for interacting with a computer using utterances, speech processing and natural language processing. Generically, the system comprises a speech processor for searching a first grammar file for a matching phrase for the utterance, and for searching a second grammar file for the matching phrase if the matching phrase is not found in the first grammar file. The system also 15 includes a natural language processor for searching a database for a matching entry for the matching phrase; and an application interface for performing an action associated with the matching entry if the matching entry is found in the database.

20 In the preferred embodiment, the natural language processor updates at least one of the database, the first grammar file and the second grammar file with the matching phrase if the matching entry is not found in the database.

25 The first grammar file is a context-specific grammar file. A context-specific grammar file is one which contains words and phrases that are highly relevant to a specific subject. The second grammar file is a general grammar file. A general grammar file is one which contains words and phrases which do not need to be interpreted in light of a context. That is to say, the words and phrases in the general grammar file do not belong to any parent context. By searching the context-specific grammar file before searching the general grammar file, the present invention allows the user to communicate with the computer using a more conversational style, wherein the words spoken, if found in the context specific grammar file, are interpreted in light of the subject matter most recently discussed.

30 In a further aspect of the present invention, the speech processor searches a dictation grammar for the matching phrase if the matching phrase is not found in the general grammar file. The dictation grammar is a large vocabulary of general words and phrases. By searching the context-specific and general grammars first, it is expected that the speech

recognition time will be greatly reduced due to the context-specific and general grammars being physically smaller files than the dictation grammar.

In another aspect of the present invention, the natural language processor replaces at least one word in the matching phrase prior to searching the database. This may be accomplished by a variable replacer in the natural language processor for substituting a wildcard for the at least one word in the matching phrase. By substituting wildcards for certain words (called "word-variables") in the phrase, the number of entries in the database can be significantly reduced. Additionally, a pronoun substituter in the natural language processor may substitute a proper name for pronouns in the matching phrase, allowing user-specific facts to be stored in the database.

In another aspect of the present invention, a string formatter text formats the matching phrase prior to searching the database. Also, a word weighter weights individual words in the matching phrase according to a relative significance of the individual words prior to searching the database. These steps allow for faster, more accurate searching of the database.

A search engine in the natural language processor generates a confidence value for the matching entry. The natural language processor compares the confidence value with a threshold value. A boolean tester determines whether a required number of words from the matching phrase are present in the matching entry. This boolean testing serves as a verification of the results returned by the search engine.

In order to clear up ambiguities, the natural language processor prompts the user whether the matching entry is a correct interpretation of the utterance if the required number of words from the matching phrase are not present in the matching entry. The natural language processor also prompts the user for additional information if the matching entry is not a correct interpretation of the utterance. At least one of the database, the first grammar file and the second grammar file are updated with the additional information. In this way, the present invention adaptively "learns" the meaning of additional utterances, thereby enhancing the efficiency of the user interface.

The speech processor will enable and search a context-specific grammar associated with the matching entry for a subsequent matching phrase for a subsequent utterance. This ensures that the most relevant words and phrases will be searched first, thereby decreasing speech recognition times.

Generically, the invention includes a method for updating a computer for voice interaction with a network object, such as a web-page. Initially, a network object table, which associates with the network object with the voice interaction system, is transferred to the computer over a network. The location of the network object table can be imbedded 5 within the network object, at a specific internet web-site, or at consolidated location that stores network object tables for multiple network objects. The network object table is searched for an entry matching the network object. The entry matching the network object may result in an action being performed, such as text speech being voiced through a speaker, a context-specific grammar file being used, or a natural language processor database being used. The network object table may be part of a dialog definition file. Dialog definition 10 files may also include a context-specific grammar, entries for a natural language processor database, or both.

In another aspect of the present invention, a network interface transfers a dialog definition file from over the network. The dialog definition file contains a network object table. A data processor searches the network object table for a table entry that matches the 15 network object. Once this matching table entry is found, an application interface performs an action specified by the matching entry.

In another aspect of the present invention, the dialog definition file associated with a network is located, and then read. The dialog definition file could be read from a variety of 20 locations, such as a web-site, storage media, or a location that stores dialog definition files for multiple network objects. A network object table, contained within the dialog definition file, is searched to find a table entry matching the network object. The matching entry defines an action associated with the network object, and the action is then performed by the system. In addition to a network object table, the dialog definition file may contain a 25 context-specific grammar, entries for a natural language processor database or both.

BRIEF DESCRIPTION OF THE DRAWINGS

The features, objects and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the 30 drawings in which like reference characters identify correspondingly throughout and wherein:

FIG. 1 is a functional block diagram of an exemplary computer system for use with the present invention;

FIG. 2 is an expanded functional block diagram of the CPU 102 and storage medium 108 of the computer system of FIG. 1 of the present invention;

FIGS. 3A-3D are a flowchart of the method of providing interactive speech recognition and natural language processing to a computer;

5 FIG. 4 is a diagram of selected columns of an exemplary natural language processing (NLP) database of the present invention;

FIG. 5 is a diagram of an exemplary Database Definition File (DDF) according to the present invention;

10 FIG. 6 is a diagram of selected columns of an exemplary network object table of the present invention;

FIGS. 7A-7C are a flowchart of the method of the present invention, illustrating the linking of interactive speech recognition and natural language processing to a networked object, such as a web-page; and

15 FIG. 8 is a diagram depicting a computer system connecting to other computers, storage media, and web-sites via the Internet.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention will now be disclosed with reference to a functional block diagram of an exemplary computer system 100 of FIG. 1. In FIG. 1, computer system 100 includes a central processing unit (CPU) 102. The CPU 102 may be any general purpose microprocessor or microcontroller as is known in the art, appropriately programmed to perform the method described herein with reference to FIGS. 3A-3D. The software for programming the CPU can be found at storage medium 108 or alternatively from another location across a computer network. For example, CPU 102 may be a conventional microprocessor such as the Pentium II processor manufactured by Intel Corporation or the like.

CPU 102 communicates with a plurality of peripheral equipment, including a display 104, manual input 106, storage medium 108, microphone 110, speaker 112, data input port 114 and network interface 116. Display 104 may be a visual display such as a CRT, LCD screen, touch-sensitive screen, or other monitors as are known in the art for visually displaying images and text to a user. Manual input 106 may be a conventional keyboard, keypad, mouse, trackball, or other input device as is known in the art for the manual input of data. Storage medium 108 may be a conventional read/write memory such as a magnetic

disk drive, floppy disk drive, CD-ROM drive, silicon memory or other memory device as is known in the art for storing and retrieving data. Significantly, storage medium 108 may be remotely located from CPU 102, and be connected to CPU 102 via a network such as a local area network (LAN), or a wide area network (WAN), or the Internet. Microphone 110 may 5 be any suitable microphone as is known in the art for providing audio signals to CPU 102. Speaker 112 may be any suitable speaker as is known in the art for reproducing audio signals from CPU 102. It is understood that microphone 110 and speaker 112 may include appropriate digital-to-analog and analog-to-digital conversion circuitry as appropriate. Data input port 114 may be any data port as is known in the art for interfacing with an external 10 accessory using a data protocol such as RS-232, Universal Serial Bus, or the like. Network interface 116 may be any interface as known in the art for communicating or transferring files across a computer network, examples of such networks include TCP/IP, ethernet, or token ring networks. In addition, on some systems, a network interface 116 may consist of a modem connected to the data input port 114.

15 Thus, FIG. 1 illustrates the functional elements of a computer system 100. Each of the elements of computer system 100 may be suitable off-the-shelf components as described above. The present invention provides a method and system for human interaction with the computer system 100 using speech.

20 As shown in FIG. 8, the computer system 100 may be connected to the Internet 700, a collection of computer networks. To connect to the Internet 700, computer system 100 may use a network interface 116, a modem connected to the data input port 114, or any other method known in the art. Web-sites 710, other computers 720, and storage media 108 may also be connected to the Internet through such methods known in the art.

25 Turning now to FIG. 2, FIG. 2 illustrates an expanded functional block diagram of CPU 102 and storage medium 108. It is understood that the functional elements of FIG. 2 may be embodied entirely in software or hardware or both. In the case of a software embodiment, the software may be found at storage medium 108 or at an alternate location across a computer network. CPU 102 includes speech recognition processor 200, data processor 201, natural language processor 202, and application interface 220. The data processor 201 interfaces with the display 104, storage medium 108, microphone 110, speaker 112, data input port 114, and network interface 116. The data processor 201 allows 30 the CPU to locate and read data from these sources. Natural language processor 202 further includes variable replacer 204, string formatter 206, word weighter 208, boolean tester 210,

pronoun replacer **211**, and search engine **213**. Storage medium **108** includes a plurality of context-specific grammar files **212**, general grammar file **214**, dictation grammar **216**, and natural language processor (NLP) database **218**. In the preferred embodiment, the grammar files **212**, **214**, and **216** are Bakus-Naur Form (BNF) files which describe the structure of the language spoken by the user. BNF files are well-known in the art for describing the structure of language, and details of BNF files will therefore not be discussed herein. One advantage of BNF files is that hierarchical tree-like structures may be used to describe phrases or word sequences, without the need to explicitly recite all combinations of these word sequences. Thus, the use of BNF files in the preferred embodiment minimizes the physical sizes of the files **212**, **214**, and **216** in the storage medium **108**, increasing the speed at which these files can be enabled and searched as described below. However, in alternate embodiments, other file structures are used.

The operation and interaction of these functional elements of FIG. 2 will be described with reference to the flowchart of FIGS. 3A-3D. In FIG. 3A, the flow begins at block **300** with the providing of an utterance to speech processor **200**. An utterance is a series of sounds having a beginning and an end, and may include one or more spoken words. Microphone **110** which captures spoken words may perform the step of block **300**. Alternately, the utterance may be provided to the speech processor **200** over data input port **114**, or from storage medium **108**. Preferably, the utterance is in a digital format such as the well-known ".wav" audio file format.

The flow proceeds to decision **302** where the speech processor **200** determines whether one of the context-specific grammars **212** has been enabled. If the context-specific grammars **212** are enabled, the context-specific grammars **212** are searched at block **304**. In the preferred embodiment, the context-specific grammars **212** are BNF files which contain words and phrases which are related to a parent context. In general, a context is a subject area. For example, in one embodiment of the present invention applicable to personal computers, examples of contexts may be "news", or "weather", or "stocks". In such a case, the context-specific grammars **212** would each contain commands, control words, descriptors, qualifiers, or parameters that correspond to a different one of these contexts. The use of contexts provides a hierachal structure for types of information. Contexts and their use will be described further below with reference to the NLP database **218**.

If a context-specific grammar **212** has been enabled, the context-specific grammar **212** is searched for a match to the utterance provided at block **300**. However, if a context-

specific grammar **212** has not been enabled, the flow proceeds to block **308** where the general grammar **214** is enabled.

In the preferred embodiment, the general grammar **214** is a BNF file which contains words and phrases which do not, themselves, belong to a parent context, but may have an associated context for which a context-specific grammar file **212** exists. In other words, the words and phrases in the general grammar **214** may be at the root of the hierachal context structure. For example, in one embodiment applicable to personal computers, the general grammar **214** would contain commands and control phrases.

In block **310**, the general grammar **214** is searched for a matching word or phrase for the utterance provided at block **300**. A decision is made, depending on whether the match is found, at block **312**. If a match is not found, then the dictation grammar **216** is enabled at block **314**. In the preferred embodiment, the dictation grammar **216** is a BNF file that contains a list of words that do not, themselves, have either a parent context or an associated context. For example, in one embodiment applicable to a personal computer, the dictation grammar **216** contains a relatively large list of general words similar to a general dictionary.

In block **316** the dictation grammar is searched for matching words for each word of the utterance provided at block **300**. At decision block **318**, if no matching words are found, a visual error message is optionally displayed at the display **104** or an audible error message is optionally reproduced through speaker **112**, at block **320**. The process ends until another utterance is provided to the speech processor **200** at block **300**.

Thus, as can be seen from the above description, when an utterance is provided to the speech processor **200**, the enabled context-specific grammar **212**, if any, is first searched. If there are no matches in the enabled context-specific grammar **212**, then the general grammar **214** is enabled and searched. If there are no matches in the general grammar **214**, then the dictation grammar **316** is enabled and searched.

In the preferred embodiment, when the speech recognition processor **200** is searching either the context-specific grammar **212** or the general grammar **214**, it is said to be in the "command and control" mode. In this mode, the speech recognition processor **200** compares the entire utterance as a whole to the entries in the grammar. By contrast, when the speech recognition processor **200** is searching the dictation grammar, it is said to be in the "dictation" mode. In this mode, the speech recognition processor **200** compares the utterance to the entries in the dictation grammar **216** one word at a time. It is expected that

searching for a match for an entire utterance in the command and control mode will generally be faster than searching for one word at a time in dictation mode.

It is further expected that any individual context-specific grammar **212** will be smaller in size (i.e., fewer total words and phrases) than the general grammar **214**, which in turn will be smaller in size than the dictation grammar **216**. Thus, by searching any enabled context-specific grammar **212** first, it is likely that a match, if any, will be found more quickly, due at least in part to the smaller file size. Likewise, by searching the general grammar **214** before the dictation grammar **216**, it is likely that a match, if any, will be found more quickly.

10 Additionally, as will be explained further below with regard to the ability of the present invention to adaptively add to both the context-specific grammar **212** and the general grammar **214**, they will contain the most common utterances. As such, it is expected that a match is more likely to be found quickly in the context-specific grammar **212** or the general grammar **214** than in the dictation grammar **216**.

15 Finally, as will be explained further below, the words and phrases in the enabled context-specific grammar **212** are more likely to be uttered by the user because they are words that are highly relevant to the subject matter about which the user was most recently speaking. This also allows the user to speak in a more conversational style, using sentence fragments, with the meaning of his words being interpreted according to the enabled context-
20 specific grammar **212**.

By searching in the above-described sequence, the present invention may search more efficiently than if the searching were to occur one entry at a time in a single, large list of all expected words and phrases.

Referring back to decision **306**, if a match is found in the context-specific grammar **212**, then there are three possible next steps shown in FIG. 3A. For each matching entry in the enabled context-specific grammar **212**, there may be an associated action to be taken by the speech recognition processor **200**. Block **322** shows that one action may be to direct application interface **220** to take some action with respect to a separate software application or entity. For example, application interface **220** may use the Speech Application Programming Interface (SAPI) standard by Microsoft to communicate with an external application. The external application may be directed, for example, to access a particular Internet web site URL or to speak a particular phrase by converting text to speech. Other
25

actions may be taken as will be discussed further below with reference to the NLP database 218 of FIG. 4.

Block 324 shows that another action may be to access a row in the natural language processing (NLP) database 218 directly, thereby bypassing the natural language processing steps described further below. Block 326 shows that another action may be to prepend a word or phrase for the enabled context to the matching word or phrase found in the context-specific grammar 306. For example, if the enabled context were "movies" and the matching utterance were "8 o'clock", the word "movies" would be prepended to the phrase "8 o'clock" to form the phrase "movies at 8 o'clock".

Likewise, if a match is found in the general grammar 214, then the flow may proceed to block 322 where the application interface 220 is directed to take an action as described above, or to block 324 where a row in the NLP database is directly accessed. However, if a match is found in the general grammar 214, no prepending of a context occurs because, as stated above, the entries in the general grammar 214 do not, themselves, have a parent context.

Alternatively, manually entered words may be captured, at block 301, and input into the natural language processor.

Finally, with reference to FIG. 3A, words may be entered manually via manual input 106. In this case, no speech recognition is required, and yet natural language processing of the entered words is still desired. Thus, the flow proceeds to FIG. 3B.

In FIG. 3B, at block 328, the natural language processor 202 formats the phrase for natural language processing analysis. This formatting is accomplished by string formatter 206 and may include such text processing as removing duplicate spaces between words, making all letters lower case (or upper case), expanding contractions (e.g., changing "it's" to "it is"), and the like. The purpose of this formatting step is to prepare the phrase for parsing.

The flow continues to block 330 where certain "word-variables" are replaced with an associated wildcard function by variable replacer 204 in preparation for accessing the NLP database 218. As used herein, the term "word-variables" refers to words or phrases that represent amounts, dates, times, currencies, and the like. For example, in one embodiment the phrase "what movies are playing at 8 o'clock" would be transformed at block 330 to "what movies are playing at \$time" where "\$time" is a wildcard function used to represent any time value. As another example, in one embodiment the phrase "sell IBM stock at 100 dollars" would be transformed at block 330 to "sell IBM stock at \$dollars" where "\$dollars"

is a wildcard function used to represent any dollar value. This step may be accomplished by a simple loop that searches the phrase for key tokens such as the words "dollar" or "o'clock" and replaces the word-variables with a specified wildcard function. In order to keep track of the location in the phrase where the substitution was made, an array may be used. This allows re-substitution of the original word-variable back into the phrase at the same position after the NLP database 218 has been searched.

The purpose of replacing word-variables with an associated wildcard function at block 330 is to reduce the number of entries that must be present in the NLP database 218. For example, the NLP database 218 would only contain the phrase "what movies are playing at \$time" rather than a separate entry for 8 o'clock, 9 o'clock, 10 o'clock, and so on. The NLP database 218 will be described further below.

At block 332, pronouns in the phrase are replaced with proper names by pronoun replacer 211. For example, in one embodiment the pronouns "I", "my" or "mine" would be replaced with the speaker's name. The purpose of this step is to allow user-specific facts to be stored and accessed in the NLP database 218. For example, the sentence "who are my children" would be transformed into "who are Dean's children" where "Dean" is the speaker's proper name. Again, this step may be performed in a simple loop that searches the phrase for pronouns, and replaces the pronouns found with an appropriate proper name. In order to keep track of the locations in the phrase where a substitution was made, an array may be used.

In block 334, the individual words in the phrase are weighted according to their relative "importance" or "significance" to the overall meaning of the phrase by word weighter 208. For example, in one embodiment there are three weighting factors assigned. The lowest weighting factor is assigned to words such as "a", "an", "the" and other articles. The highest weighting factor is given to words that are likely to have a significant relation to the meaning of the phrase. For example, these may include all verbs, nouns, adjectives, and proper names in the NLP database 218. A medium weighting factor is given to all other words in the phrase. The purpose of this weighting is to allow for more powerful searching of the NLP database 218.

An example of selected columns of the NLP database 218 of one embodiment is shown in FIG. 4. The NLP database 218 comprises a plurality of columns 400-410, and a plurality of rows 412A-412N. In column 400, the entries represent phrases that are "known" to the NLP database. In column 402, a number of required words for each entry in column

400 is shown. In column 404, an associated context or subcontext for each entry in column 400 is shown. In columns 408 and 410, one or more associated actions are shown for each entry in column 400. It should be noted that the NLP database 218 shown in FIG. 4 is merely a simplified example for the purpose of teaching the present invention. Other 5 embodiments may have more or fewer columns with different entries.

Referring back to FIG. 3B, at block 336, the NLP database 218 is searched for possible matches to the phrase, based on whether the entry in column 400 of the NLP database 218 contains any of the words in the phrase (or their synonyms), and the relative weights of those words. At block 338, a confidence value is generated for each of the 10 possible matching entries based on the number of occurrences of each of the words in the phrase and their relative weights. Weighted word searching of a database is well known in the art and may be performed by commercially-available search engines such as the product "dtsearch" by DT Software, Inc. of Arlington, Virginia. Likewise, searching using 15 synonyms is well known in the art and may be accomplished using such publicly-available tools such as "WordNet", developed by the Cognitive Science Laboratory of Princeton University in Princeton, New Jersey. The search engine may be an integral part of the natural language processor 202.

At decision 340, the natural language processor 202 determines whether any of the 20 possible matching entries has a confidence value greater than or equal to some predetermined minimum threshold, T. The threshold T represents the lowest acceptable confidence value for which a decision can be made as to whether the phrase matched any of the entries in the NLP database 218. If there is no possible matching entry with a confidence value greater than or equal to T, then the flow proceeds to block 342 where an optional error message is either visually displayed to the user over display 104 or audibly reproduced over speaker 112. In one embodiment, the type of error message, if any, displayed to the user 25 may depend on how many "hits" (i.e., how many matching words from the phrase) were found in the highest-confidence NLP database entry. A different type of error message would be generated if there was zero or one hits, than if there were two or more hits.

If, however, there is at least one entry in the NLP database 218 for which a 30 confidence value greater than or equal to T exists, then the flow proceeds to block 344 where the "noise" words are discarded from the phrase. The "noise" words include words which do not contribute significantly to the overall meaning of the phrase relative to the other words in the phrase. These may include articles, pronouns, conjunctions, and words of a similar

nature. "Non-noise" words would include words which contribute significantly to the overall meaning of the phrase. "Non-noise" words would include verbs, nouns, adjectives, proper names, and words of a similar nature.

The flow proceeds to FIG. 3C where the non-noise word requirement is retrieved from column 402 of the NLP database 218 for the highest-confidence matching entry at block 346. For example, if the highest-confidence matching phrase was the entry in row 412A, (e.g., "what movies are playing at \$time"), then the number of required non-noise words is 3.

At decision 348, a test is made to determine whether the number of required non-noise words from the phrase is actually present in the highest-confidence entry retrieved from the NLP database 218. This test is a verification of the accuracy of the relevance-style search performed at block 336, it being understood that an entry may generate a confidence value higher than the minimum threshold, T, without being an acceptable match for the phrase.

The nature of the test performed at decision 348 is a boolean "AND" test performed by boolean tester 210. The test determines whether each one of the non-noise words in the phrase (or its synonym) is actually present in the highest-confidence entry. If there are a sufficient number of required words actually present in the highest-confidence entry, then the flow proceeds to block 350, where the natural language processor 202 directs application interface 220 to take an associated action from column 408 or 410. It is understood that additional action columns may also be present.

For example, if the highest confidence entry was the entry in row 412A, and the boolean test of decision 348 determined that there actually were 3 non-noise words from the phrase in the entry in column 400, then the associated action in column 408 (e.g., access movie web site) would be taken. Other entries in the NLP database have other associated actions. For example, if the highest-confidence entry is that in row 412E (e.g., "what time is it"), the associated action may be for natural language processor 202 to direct a text-to-speech application (not shown) to speak the present time to the user through the speaker 112. As another example, if the highest-confidence entry is that in row 412N (e.g., "show me the news"), the first associated action may be to access a predetermined news web site on the Internet, and a second associated action may be to direct an image display application (not shown) to display images associated with the news. Different or additional actions may also be performed.

Also, if the highest-confidence entry contains the required number of non-noise words from the phrase as determined at decision 348, the natural language processor 202 instructs the speech recognition processor 200 to enable the context-specific grammar 212 for the associated context of column 404. Thus, for row 412A, context-specific grammar 212 for the context "movies" would be enabled. Thus, when the next utterance is provided to the speech recognition processor 200 in block 300 of FIG. 3A, it would search the enabled context-specific grammar 212 for "movies" before searching the general grammar 214. As previously stated, enabling the appropriate context-specific grammar 212 greatly increases the likelihood of fast, successful speech recognition, and enhances the user's ability to communicate with the computer in a conversational style.

If, however, back at decision 348, the required number of non-noise words from the phrase is not actually present in the highest-confidence entry retrieved from the NLP database 218, then the flow proceeds to block 354 where the user is prompted over display 104 or speaker 112 whether the highest-confidence entry was meant. For example, if the user uttered "How much is IBM stock selling for today", the highest-confidence entry in the NLP database 218 may be the entry in row 412B. In this case, although the relevance factor may be high, the number of required words (or their synonyms) may not be sufficient. Thus, the user would be prompted at block 354 whether he meant "what is the price of IBM stock on August 28, 1998".

The user may respond either affirmatively or negatively. If it is determined at decision 356 that the user has responded affirmatively, then the action(s) associated with the highest-confidence entry are taken at block 350, and the associated context-specific grammar 212 enabled at block 352.

If, however, it is determined at decision 356 that the user has responded negatively, then the flow proceeds to FIG. 3D where the associated context from column 404 of NLP database 218 is retrieved for the highest-confidence entry, and the user is prompted for information using a context-based interactive dialog at block 360. For example, if the user uttered "what is the price of XICOR stock today", and the highest confidence entry from the NLP database 218 was row 412B (e.g., "what is the price of IBM stock on \$date), then the user would be prompted at block 354 whether that was what he meant.

If the user responds negatively, then the context "stock" is retrieved from column 404 at block 358, and the context-based interactive dialog for the stock context is presented to the user over the display 104 and speaker 112. Such a context-based interactive dialog

may entail prompting the user for the name and stock ticker symbol of XICOR stock. The user may respond by speaking the required information. A different context-based interactive dialog may be used for each of the possible contexts. For example, the "weather" context-based interactive dialog may entail prompting the user for the name of the location (e.g., the city) about which weather information is desired. Also, the "news" context-based interactive dialog may entail prompting the user for types of articles, news source, Internet URL for the news site, or other related information.

Upon completion of the context-based interactive dialog, the NLP database 218, general grammar 214, and context-specific grammar 212 are updated to include the new information, at block 362. In this way, the next time the user asks for that information, a proper match will be found, and the appropriate action taken without prompting the user for more information. Thus, the present invention adaptively "learns" to recognize phrases uttered by the user.

In one embodiment of the present invention, one or more of the NLP database 218, context specific grammar 212, general grammar 214, and dictation grammar 216 also contain time-stamp values (not shown) associated with each entry. Each time a matching entry is used, the time-stamp value associated with that entry is updated. At periodic intervals, or when initiated by the user, the entries that have a time-stamp value before a certain date and time are removed from their respective databases/grammars. In this way, the databases/grammars may be kept to an efficient size by "purging" old or out-of-date entries. This also assists in avoiding false matches.

In one embodiment of the present invention, speech recognition and natural language processing may be used to interact with networked objects, such as World-Wide-Web ("WWW" or "web") pages that have a context-sensitive voice-based interface.

FIG. 5 illustrates an exemplary Dialog Definition File (DDF) 500 which represents information necessary to associate the speech recognition and natural language processing to an internet object, such as a text or graphics file or, in the preferred embodiment, a web-page. Although in its simplest embodiment the Dialog Definition File 500 consists of a network object table 510, the DDF may also contain additional context-specific grammar files 214 and additional entries for the natural language processing (NLP) database 218, as illustrated in FIG. 5. The preferred embodiment of the DDF 500 includes a network object table 510, a context-specific grammar file 214, and a file containing entries to the natural language processing database 218. These components may be compressed and combined

into the DDF file 500 by any method known in the art, such as through Lempel-Ziv compression. The context-specific specific grammar file 214 and the natural language processing database 218 are as described in earlier sections. The network object table 510 is a memory structure, such as a memory tree, chain or table, which associates an address of a resource with various actions, grammars, or entries in the NLP database 218.

An exemplary embodiment of the network object table 510 is illustrated in FIG. 6. FIG. 6 illustrates a memory table which may contain entry columns for: a network object 520, a Text-to-Speech (TTS) flag 522, a text speech 524, a use grammar flag 526, an append grammar flag 528, an “is yes/no?” flag, and “do yes” 532 and “do no” 534 actions. Each row in the table 540A-540E would represent the grammar and speech related to an individual network object. The exemplary embodiment of the invention would refer to network objects 520 through a Universal Resource Locator (URL). A URL is a standard method of specifying the address of any resource on the Internet that is part of the World-Wide-Web. As this standard is well-known in the art for describing the location of Internet resources, the details of URLs will therefore not be discussed herein. One advantage of URLs is that they can specify information in a large variety of network object formats, including hypertext, graphical, database and other files, in addition to a number of network object devices and communication protocols.

When combined with the text speech 524, the Text-to-Speech (TTS) flag 522 indicates whether an initial statement should be voiced over speaker 112 when the corresponding network object is transferred. For example, when transferring the web-page listed in the network object column 520 of row 540A (<http://www.conversationalsys.com>), the TTS flag 522 is marked, indicating the text speech 524, “Hello, welcome to Conversational Systems,” is to be voiced over speaker 112.

The next three flags relate to the use of grammars associated with this network object. The affirmative marking of the “use grammar” 526 or “append grammar” 528 flags indicate the presence of a content-specific grammar file 214 related to the indicated network object. The marking of the “use grammar” flag 526 indicates that the new content-specific grammar file 214 replaces the existing content-specific grammar file, and the existing file is disabled. The “append grammar” flag 528 indicates that the new content-specific grammar file should be enabled concurrently with the existing content-specific grammar file.

Lastly, the remaining columns entries relate to a “yes/no” grammar structure. If the “Is yes/no?” flag 530 is marked, then a standard “yes/no” grammar is enabled. When a

standard “yes/no” grammar is enabled, affirmative commands spoken to the computer result in the computer executing the command indicated in the “Do Yes” entry **532**. Similarly, a negative command spoken to the computer results in the computer executing the command indicated in the “Do No” entry **534**. The entries in the “Do Yes” **532** and “Do No” **534** columns may either be commands or pointers to commands imbedded in the NLP Database **218**. For example, as shown in row **540B**, the “Is Yes/No?” flag is marked. An affirmative answer, such as “yes,” given to the computer, would result in executing the corresponding command in the “Do Yes” entry **532**; in this specific case, the entry is the number “210,” a reference to the 210th command in the NLP database. An answer of “no” would result in the computer executing the 211th command in the NLP database.

Turning now to FIG. 7A, a to the method and system of providing speech and voice commands to internetworking computers, such as a computer browsing the World-Wide-Web, is illustrated. The method of FIGS. 7A-7C may be used in conjunction with the method of FIGS 3A-3D. In block **602**, a web-site URL (network object) is provided to a World-Wide-Web browser. The web browser is a program used to navigate through the Internet, and is well-known in the art. The step, at block **602**, of providing a URL to the browser, can be as simple as a user manually typing in the URL, or having a user select a “link” to the chosen web-site URL. It also may be the result of a voiced command as described earlier with reference to the action associated with each entry in the NLP database **218**. Given the URL, the computer must decide on whether it can resolve the internet address of the web-site specified within the URL, at block **604**. This resolution process is a process well-known in the art. If the computer is unable to resolve the internet address, an error message is displayed in the browser window, at block **605**, and the system is returned to its initial starting state **600**. If the internet address is resolved, the computer sends the web-site a request to for the web-page, at block **606**.

A decision is made, depending upon whether the web-site sends the web-page, at block **608**. If the web-site does not respond, or fails to send the web-page, an error message is displayed in the browser window, at block **605**, and the system is returned to its initial starting state **600**. If the web-site returns the web-page, the web-page is displayed in the browser window, at block **610**.

In decision block **612**, the computer **100** determines whether the DDF file **500** corresponding to the web-site is already present on the computer **100**. If the DDF file is present, the flow proceeds to FIG. 7C, if not the flow proceeds to FIG. 7B.

Moving to FIG. 7B, if the DDF file **500** is not present, the computer examines whether the DDF file **500** location is encoded within the web-page HyperText Markup Language (HTML) as a URL. (Note that HTML is well-known in the art, and the details of the language will therefore not be discussed herein.) Encoding DDF file location within 5 HTML code may be done either through listing the DDF file location in an initial HTML meta-tag such as:

```
<meta DDF= "http://www.conversationalsys.com/ConverseIt.ddf">
```

or directly through a scripting tag written into the variation of HTML supported by the browser,

```
10      <!--  
           <DDF= "http://www.conversationalsys.com/ConverseIt.ddf">  
         -->
```

If the DDF file location information is encoded within the web-page, the location's internet address is resolved, at block **616**, and the computer requests transfer of the DDF 15 file **500**, at block **626**.

Alternatively, if the DDF file **500** location is not encoded within the web-page, there are several alternate places that it may be stored. It may be stored in a pre-defined location at the web-site, such as a certain file location in the root directory, or at a different centralized location, such as another internet server or the storage medium **108** 20 of FIG. 1. Blocks **618** and **620** test for these possibilities. Block **618** determines whether the DDF file is located at the web-site. At this step, the computer sends query to the web-site inquiring about the presence of the DDF file **500**. If the DDF file **500** is present at the web-site, the computer requests transfer of the DDF file **500**, at block **626**. If the DDF file **500** is not located at the web-site, the computer queries the centralized location about 25 the presence of a DDF file for the web-site, at block **620**. If the DDF file is present at the web-site, the computer requests transfer of the DDF file, at block **626**. If the DDF file **500** cannot be found, the existing components of any present DDF file, such as the network object table **510**, NLP database **218** associated with the web-site and context-specific grammar **214** for any previously-visited web-site, are deactivated in block **622**. Furthermore, the web-site is treated as a non-voice-activated web-site, and only standard 30 grammar files are used, at block **624**. Standard grammar files are the grammar files existing on the system excluding any grammars associated with the content-specific grammar file associated with the network object.

If the DDF file **500** is requested at block **626**, and its transfer is unsuccessful, any existing components of any present DDF file **500** are deactivated, at block **622**, and the web-site is treated as a non-voice-activated web-site, and only standard grammar files are used, at block **624**.

5 If the DDF file **500** is requested at block **626** and its transfer is successful at block **628**, it replaces any prior DDF file, at block **630**. Any components of the DDF file **500**, such as the network object table **510**, context-specific-grammar files **214**, and NLP database **218** are extracted at block **632**. A similar technique may be used for obtaining the software necessary to implement the method illustrated in FIGS. 3A-3D, comprising 10 the functional elements of FIG. 2.

The flow moves to FIG. 7C. The network object table **510** is read into memory by the computer in block **634**. If the web-page URL is present in the site network object table **510**, as determined by block **636**, it will be represented by a row **540A-540E** of the table, as shown in FIG. 6. Each row of the network object table represents the speech-interactions available to a user for that particular web-page. If no row corresponding 15 to the web-page exists, then no-speech interaction exists for the web page, and processing ends.

If the web-page URL is present in the site network object table **510**, as determined by block **636**, the computer checks if the TTS flag **522** is marked, to determine whether a 20 text speech **524** is associated with the web-page, at block **638**. If there is a text speech **524**, it is voiced at block **640**, and flow continues. If there is a context-specific grammar file associated with the web-page, as determined by decision block **642**, it is enabled at block **644**, and then the NLP database **218** is enabled at block **646**. If no context-specific grammar file is associated with the web-page, only the NLP database **218** is enabled at 25 block **646**. Once the NLP database is enabled **646**, the system behaves as FIG. 3A-3C, as described above.

In summary, the present invention provides a method and system for a networked interactive user-interface for a computer. By the use of context-specific grammars that are tied to internet-objects through a Dialog Definition File, the present invention decreases 30 speech recognition time and increases the user's ability to communicate with internet objects, such as web-pages, in a conversational style. Furthermore, by the use of adaptive updating of the various grammars and the NLP database, the present invention further increases interactive efficiency.

The previous description of the preferred embodiments is provided to enable any person skilled in the art to make or use the present invention. The various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without the use of inventive faculty. Thus, the present invention is not intended to be limited to the embodiments shown herein, but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

I CLAIM:

CLAIMS

1. A method for updating a computer for voice interaction with a network object, the method comprising the steps of:

5 transferring a network object table associated with the network object over a network;

searching the network object table for a matching entry for the network object; and

performing an action associated with said matching entry if said matching entry is found in the network object table.

- 10 2. The method of claim 1 wherein the network object table is read from a web-site.

3. The method of claim 1 wherein the network object table is read from a location that stores network object tables for multiple network objects.

4. The method of claim 1 wherein the network object table is included in a dialog definition file which also includes a context-specific grammar.

- 15 5. The method of claim 1 wherein the network object table is included in a dialog definition file which also includes entries for a natural language processor database.

6. The method of claim 1 wherein the network object table is included in a dialog definition file which also includes a context-specific grammar and entries for a natural language processor database.

- 20 7. A system for updating a computer for voice interaction with a network object, the system comprising:

a network interface for transferring a dialog definition file associated with a network object, wherein the dialog definition file contains a network object table;

25 a data processor for searching the network object table for a matching entry for the network object; and

an application interface for performing an action associated with said matching entry if said matching entry is found in the network object table.

8. A method for updating a computer for voice interaction with a network object, the method comprising the steps of:

30 locating a dialog definition file associated with a network object, wherein the dialog definition file contains a network object table;

reading the dialog definition file;

searching the network object table for a matching entry for the network object; and

performing an action associated with said matching entry if said matching entry is found in the network object table.

9. The method of claim 8 wherein the dialog definition file is read from a web-site.

10. The method of claim 8 wherein the dialog definition file is read from a location that stores dialog definition files for multiple network objects.

5 11. The method of claim 8 wherein the dialog definition file is read from storage media.

12. The method of claim 8 wherein the dialog definition file includes a context-specific grammar.

13. The method of claim 8 wherein the dialog definition file includes entries for a natural language processor database.

10 14. The method of claim 8 wherein the dialog definition file includes a context-specific grammar and entries for a natural language processor database.

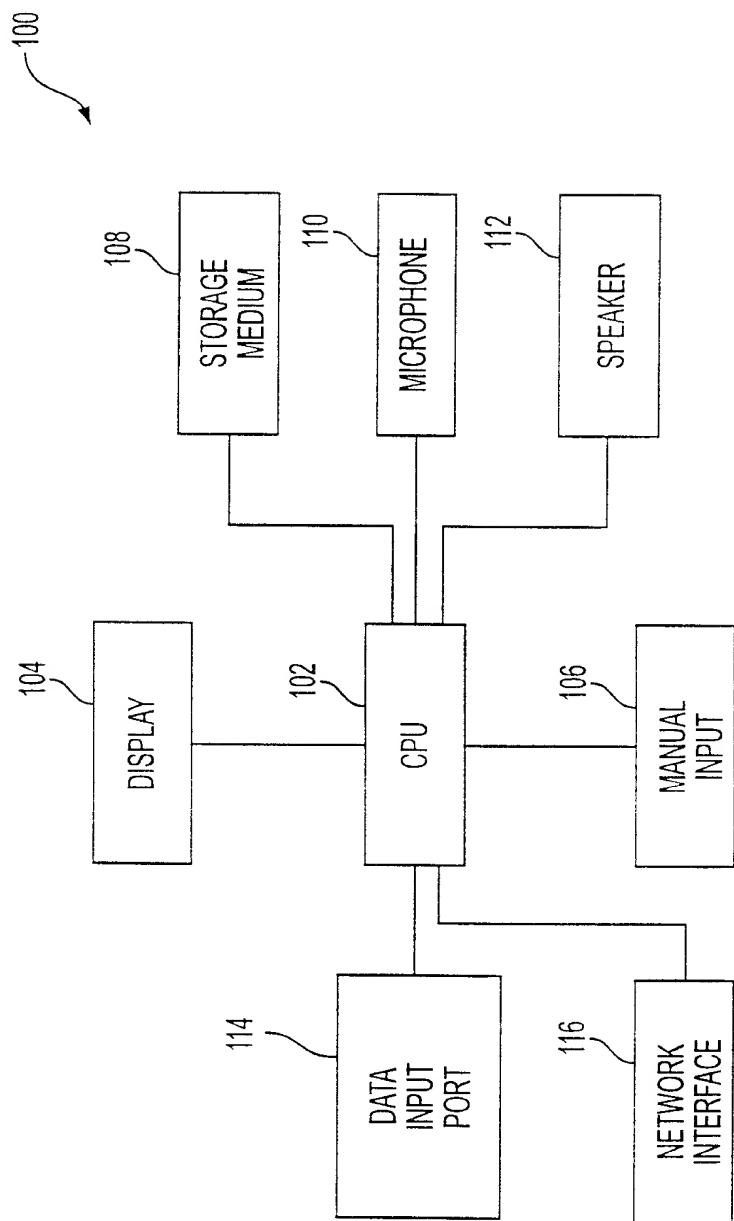
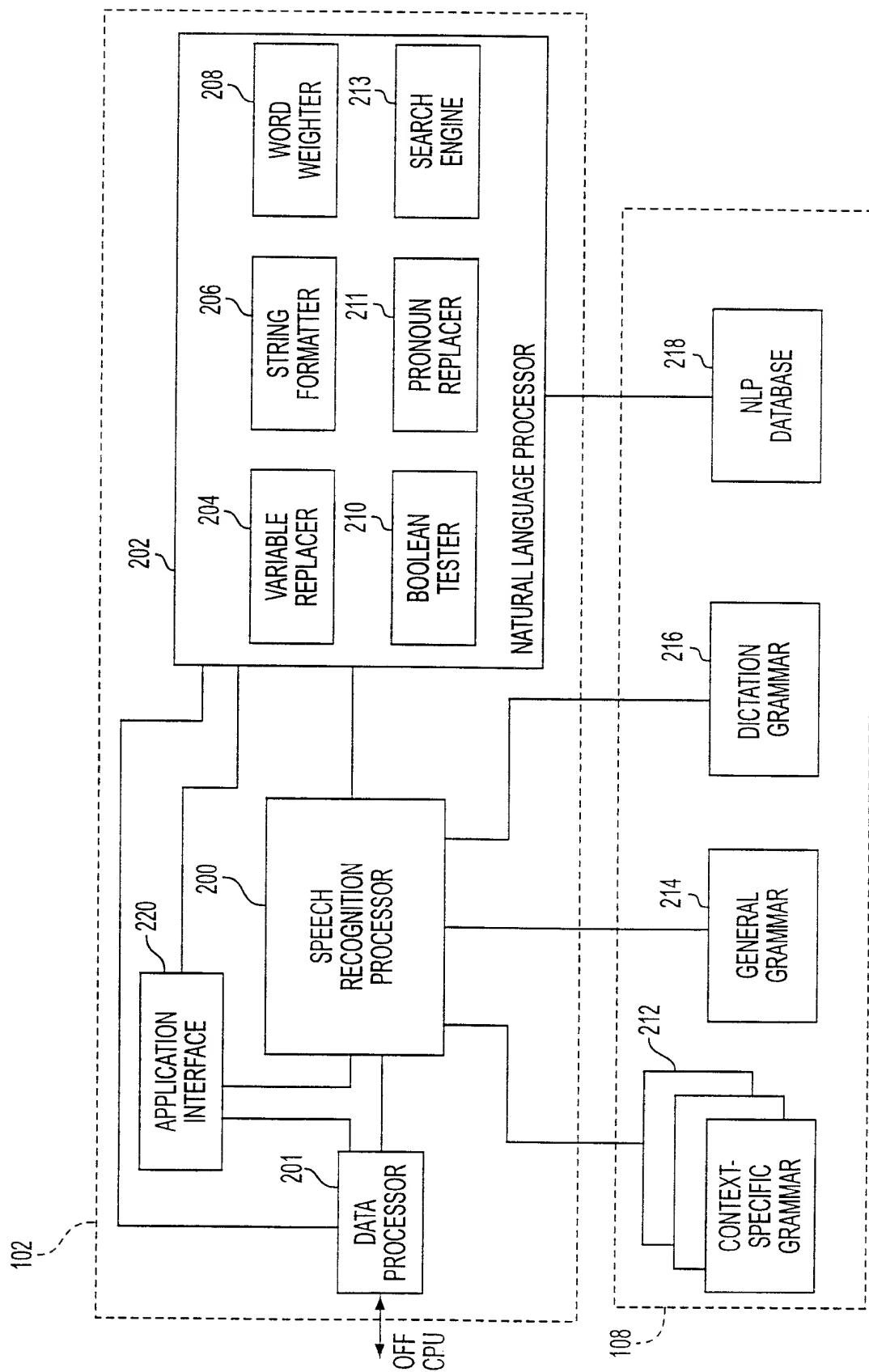


FIG. 1



SUBSTITUTE SHEET (RULE 26)

FIG. 2

3/13

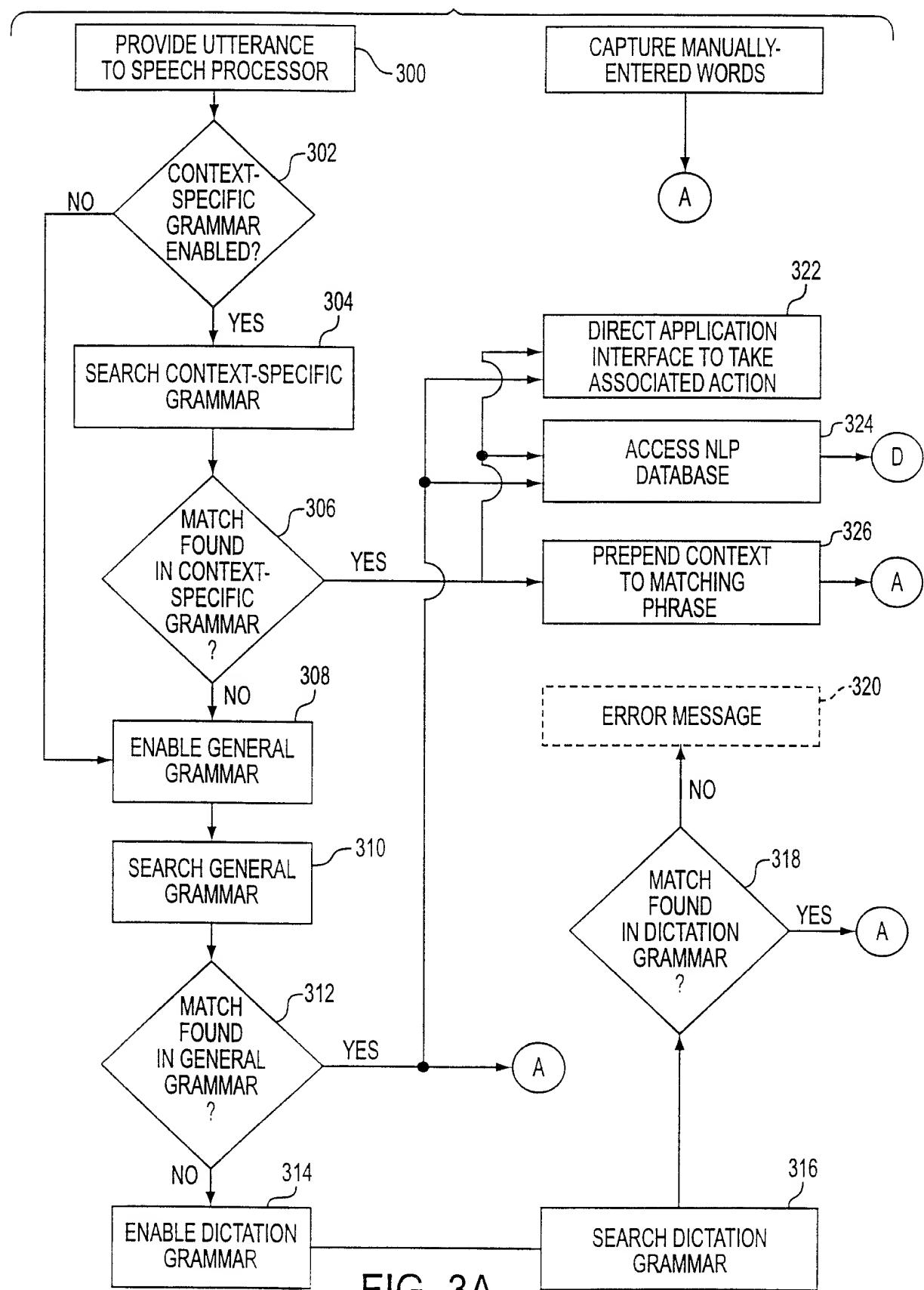


FIG. 3A

4/13

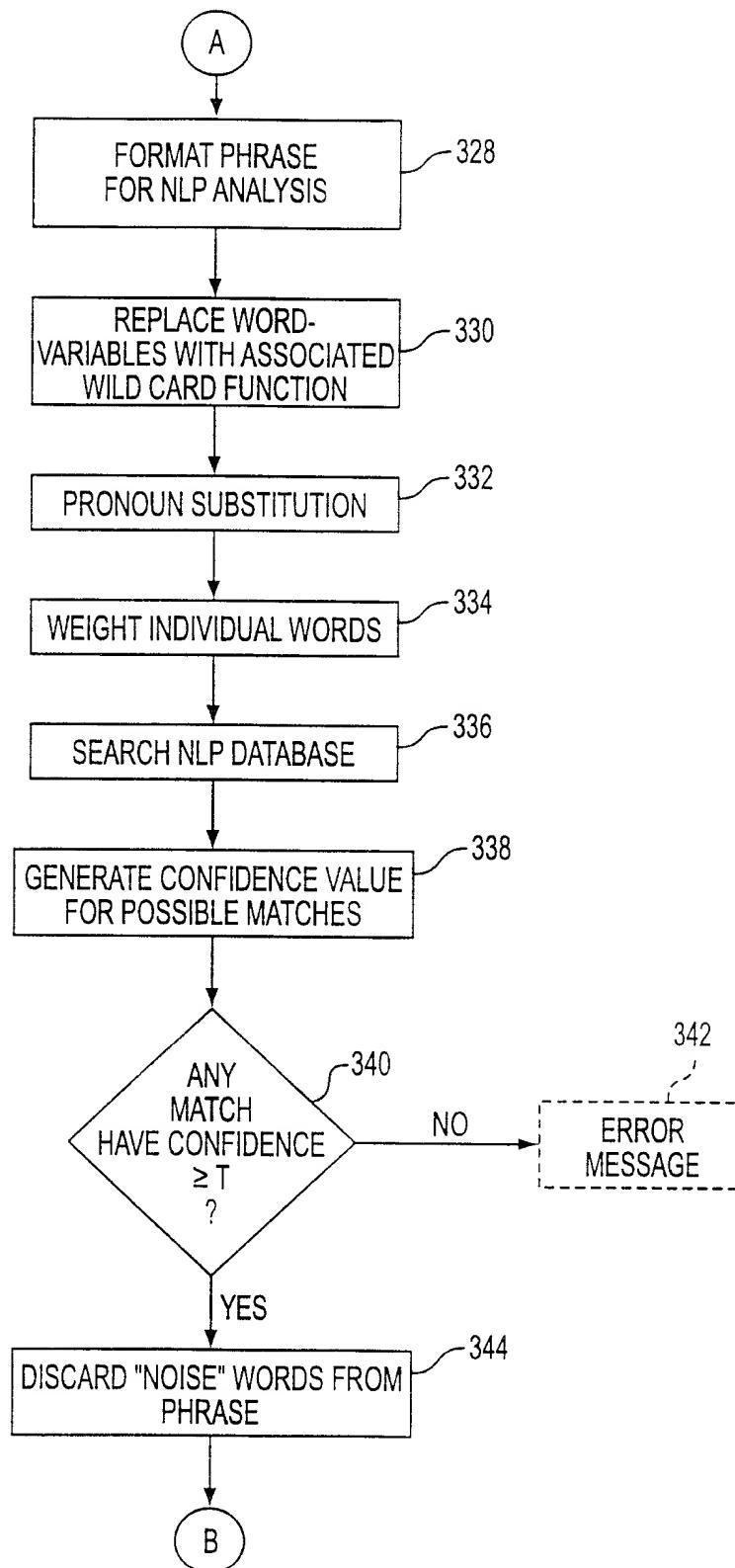


FIG. 3B

5/13

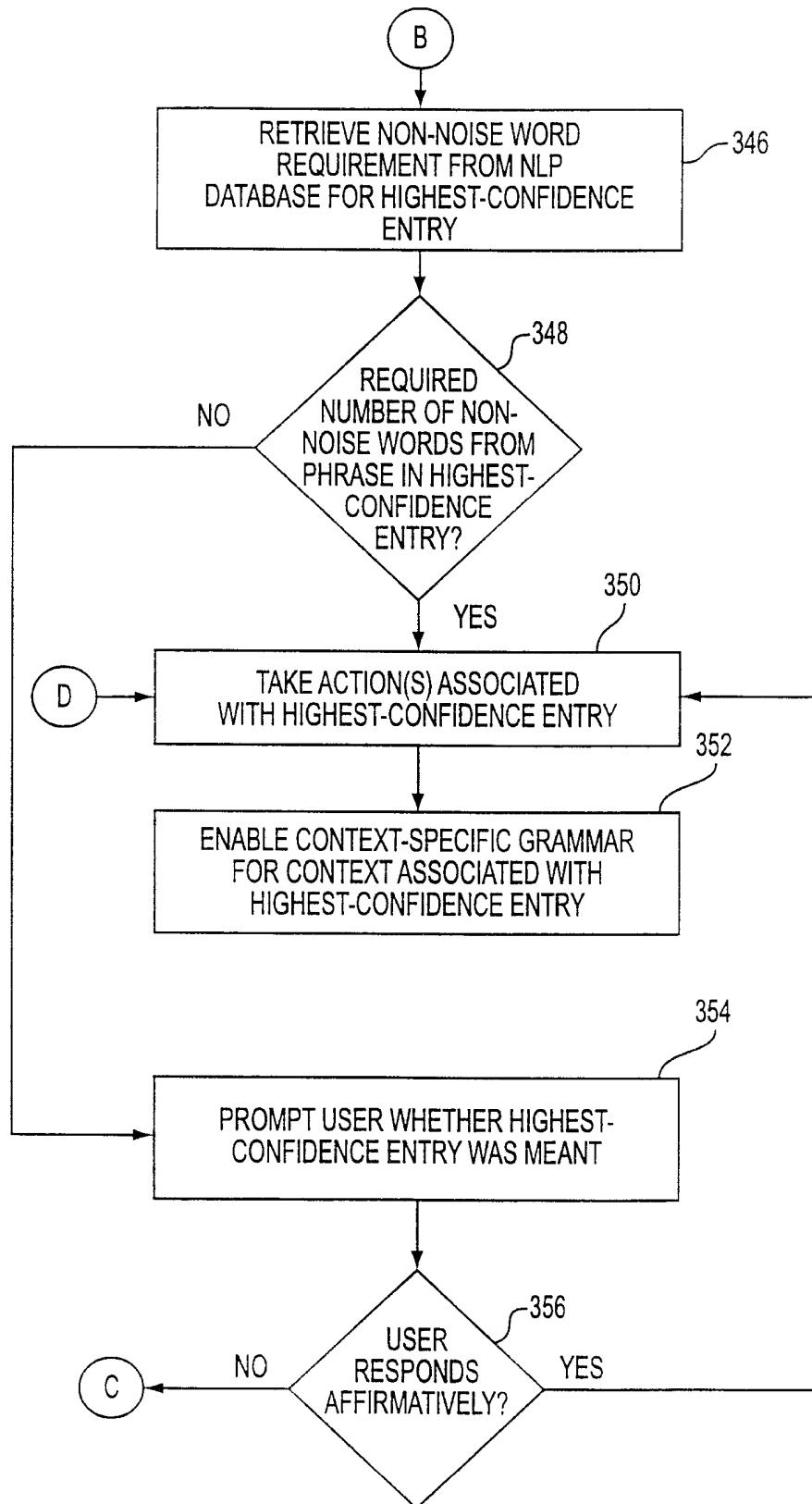


FIG. 3C

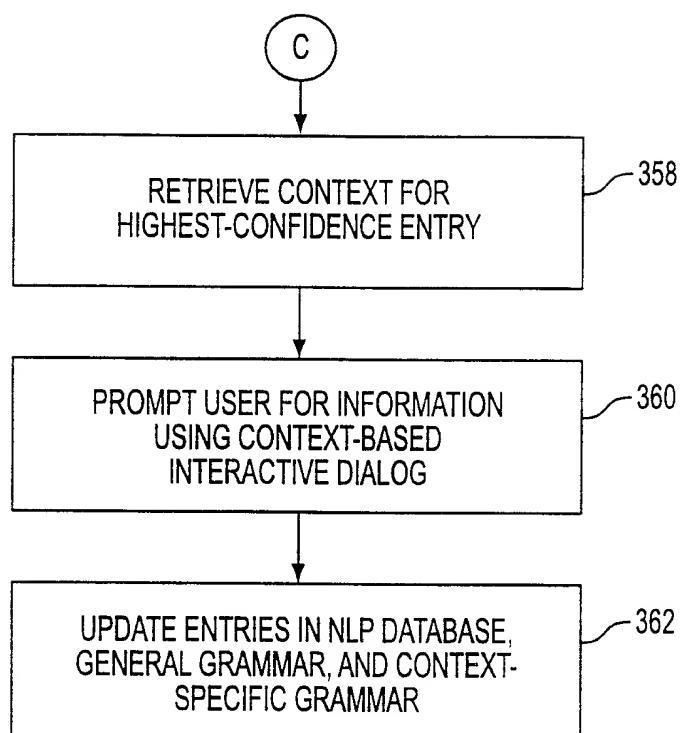


FIG. 3D

7/13

PHRASE	REQUIRED WORDS	CONTEXT/SUBCONTEXT	ACTION 1	ACTION 2
WHAT MOVIES ARE PLAYING AT \$ TIME	3	MOVIES	ACCESS MOVIE WEB SITE	N/A
WHAT IS THE PRICE OF IBM STOCK ON \$ DATE	4	STOCKS	ACCESS STOCK PRICE WEB SITE	N/A
SELL IBM STOCK AT \$ DOLLARS	4	STOCKS	ACCESS STOCK TRADE WEB SITE	N/A
WHAT IS THE WEATHER IN \$ LOCATION	3	WEATHER	ACCESS WEATHER WEB SITE	N/A
WHAT TIME IS IT	2	TIME	N/A	TEXT-TO-SPEECH OF TIME
SHOW ME THE NEWS	2	NEWS	ACCESS NEWS WEB SITE	DISPLAY IMAGES

Diagram annotations:

- Group 1: 412A (WHAT MOVIES ARE PLAYING AT \$ TIME), 412B (WHAT IS THE PRICE OF IBM STOCK ON \$ DATE), 412C (SELL IBM STOCK AT \$ DOLLARS), 412D (WHAT IS THE WEATHER IN \$ LOCATION), 412E (WHAT TIME IS IT), 412N (SHOW ME THE NEWS). These are grouped by a bracket labeled 218.
- Group 2: 400 (WHAT MOVIES ARE PLAYING AT \$ TIME), 402 (WHAT IS THE PRICE OF IBM STOCK ON \$ DATE), 404 (SELL IBM STOCK AT \$ DOLLARS), 408 (WHAT IS THE WEATHER IN \$ LOCATION), 410 (WHAT TIME IS IT). These are grouped by a bracket labeled 400.
- Group 3: 404 (WHAT IS THE PRICE OF IBM STOCK ON \$ DATE), 408 (WHAT IS THE WEATHER IN \$ LOCATION). These are grouped by a bracket labeled 404.
- Group 4: 410 (WHAT TIME IS IT). This is grouped by a bracket labeled 410.

FIG. 4

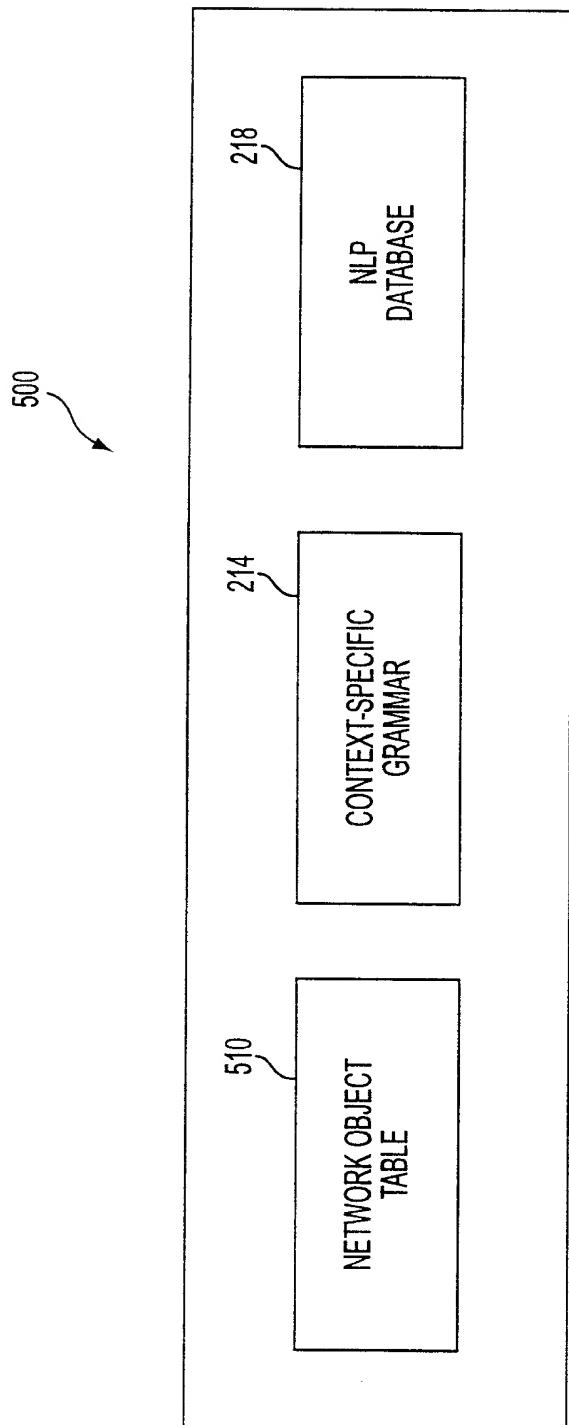


FIG. 5

9/13

NETWORK OBJECT	TTS FLAG	TEXT SPEECH	USE GRAMMAR	APPEND GRAMMAR	IS YESNO?	DO YES	DO NO
HTTP://WWW.CONVERSATIONALSYS.COM	X	"HELLO, WELCOME TO CONVERSATIONAL SYSTEMS."		X			
HTTP://WWW.CONVERSATIONALSYS.COM/NEWS	X	"WOULD YOU LIKE TO LEARN ABOUT NEW PRODUCTS?"			X	210	211
HTTP://WWW.CONVERSATIONALSYS.COM/PRODUCTS	X	"ALL NATURAL LANGUAGE PRODUCTS ARE..."	X				
HTTP://WWW.CONVERSATIONALSYS.COM/PRODUCTS/IAVOICE	X	"VIA VOICE IS..."					
HTTP://WWW.CONVERSATIONALSYS.COM/SEARCH					X	254	213

FIG. 6

10/13

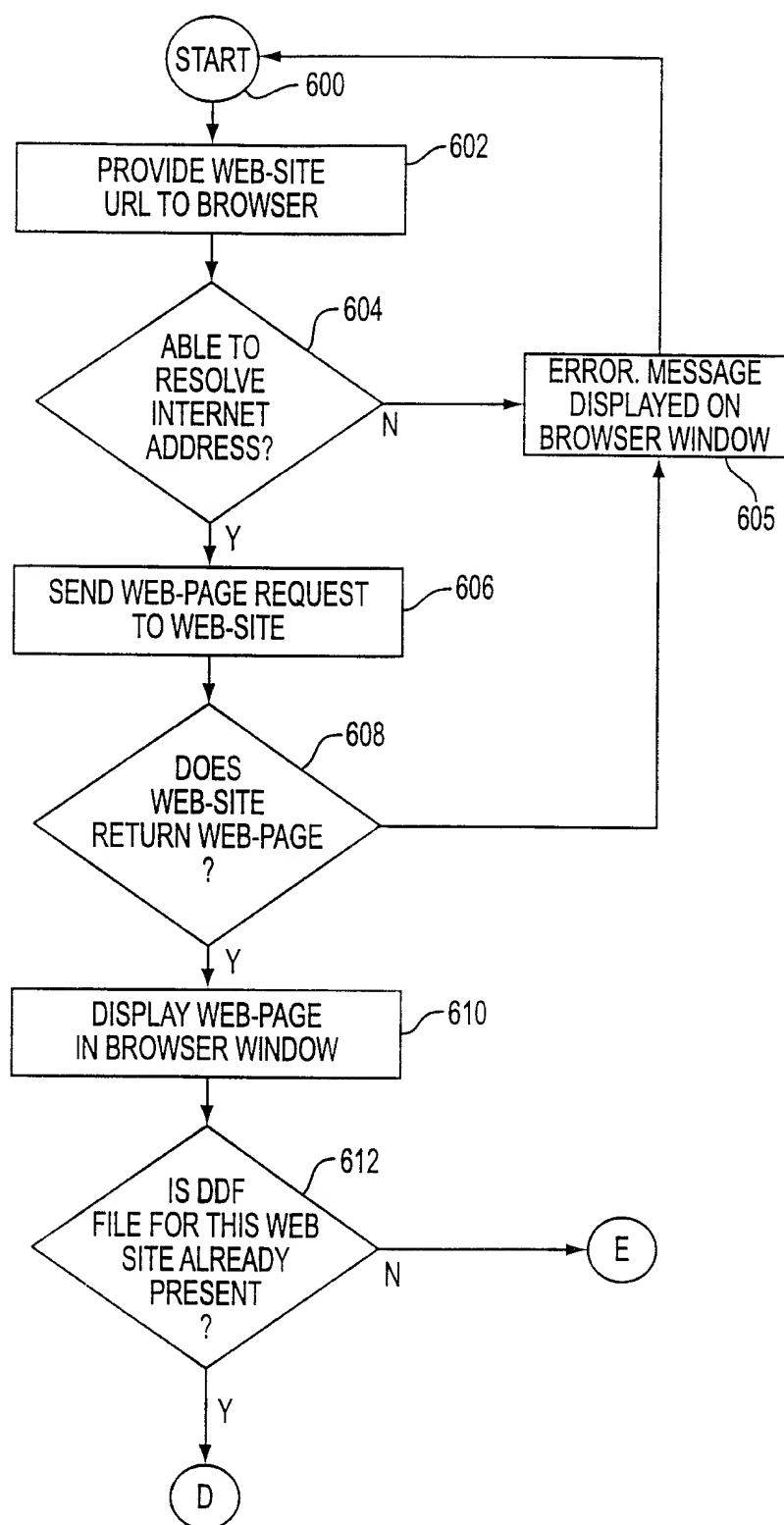


FIG. 7A

11/13

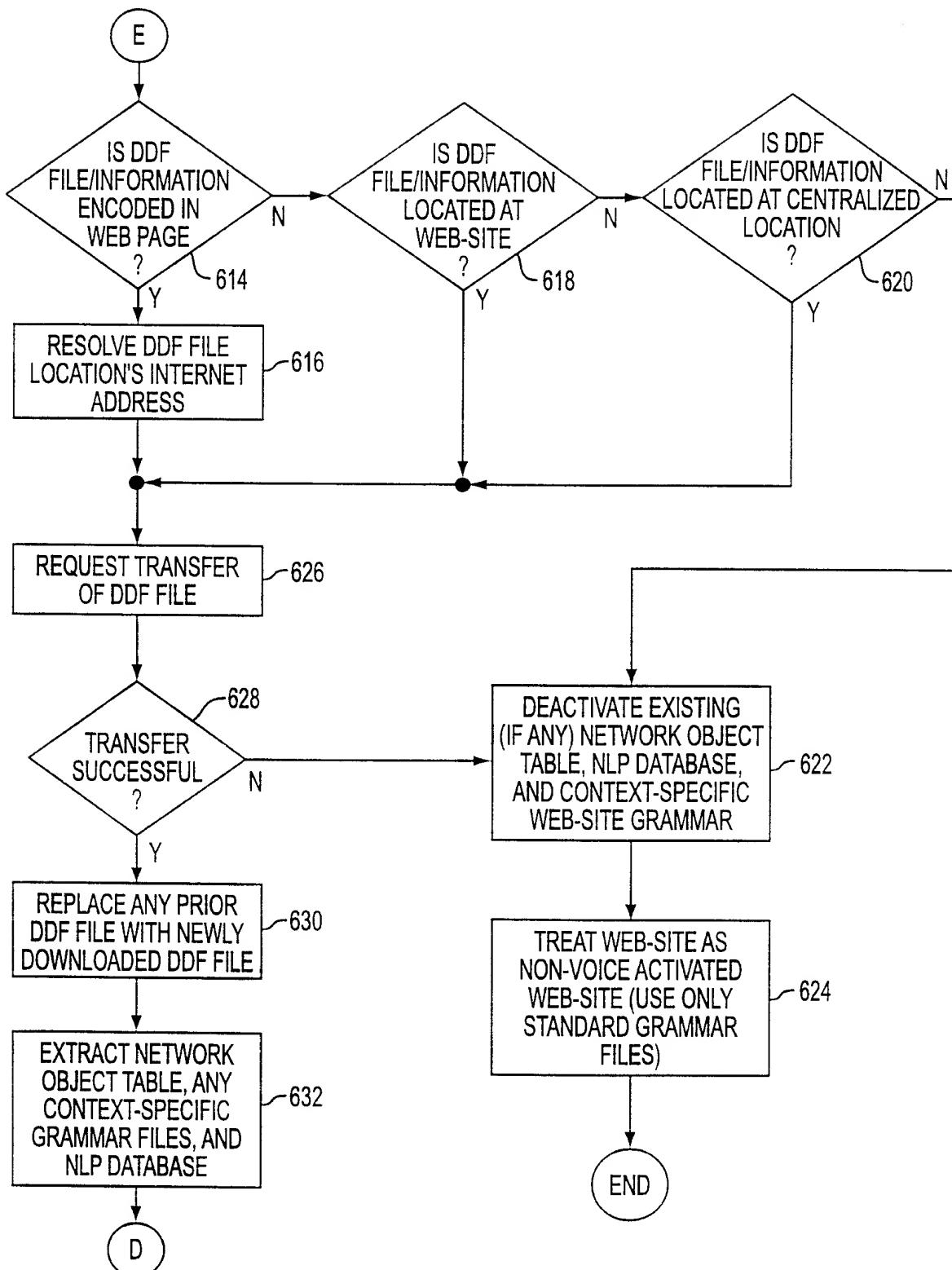


FIG. 7B

12/13

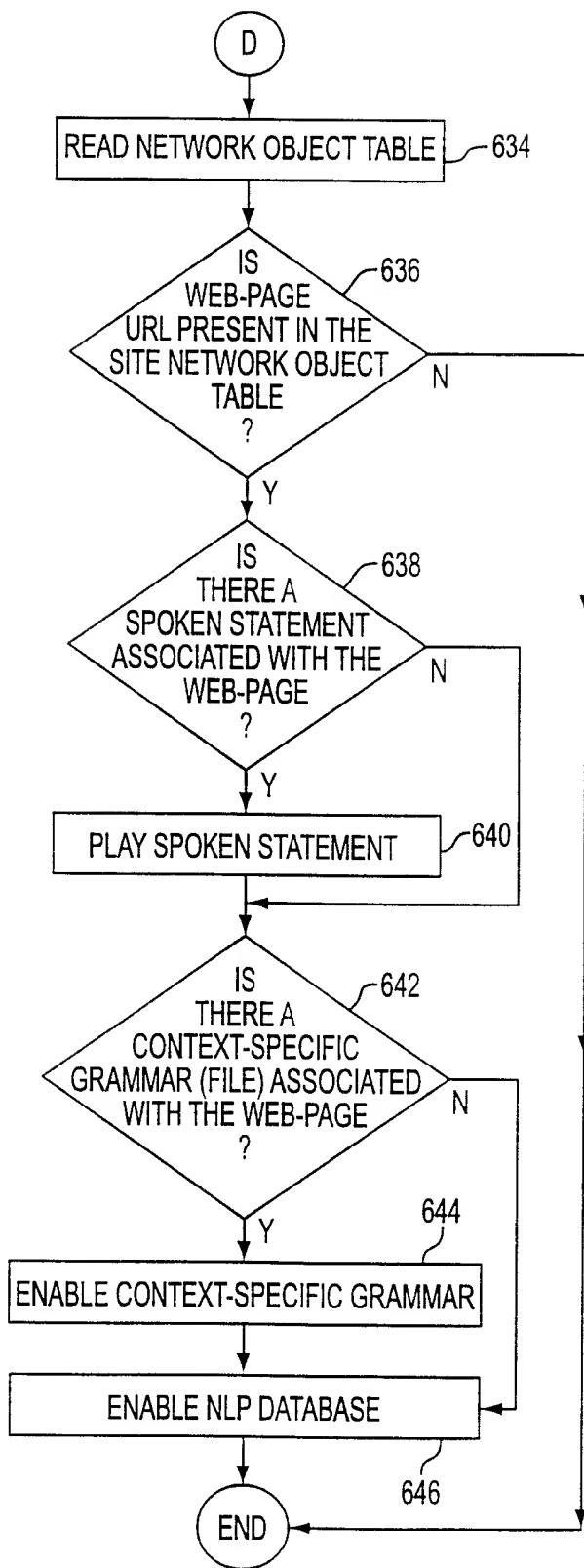


FIG. 7C
SUBSTITUTE SHEET (RULE 26)

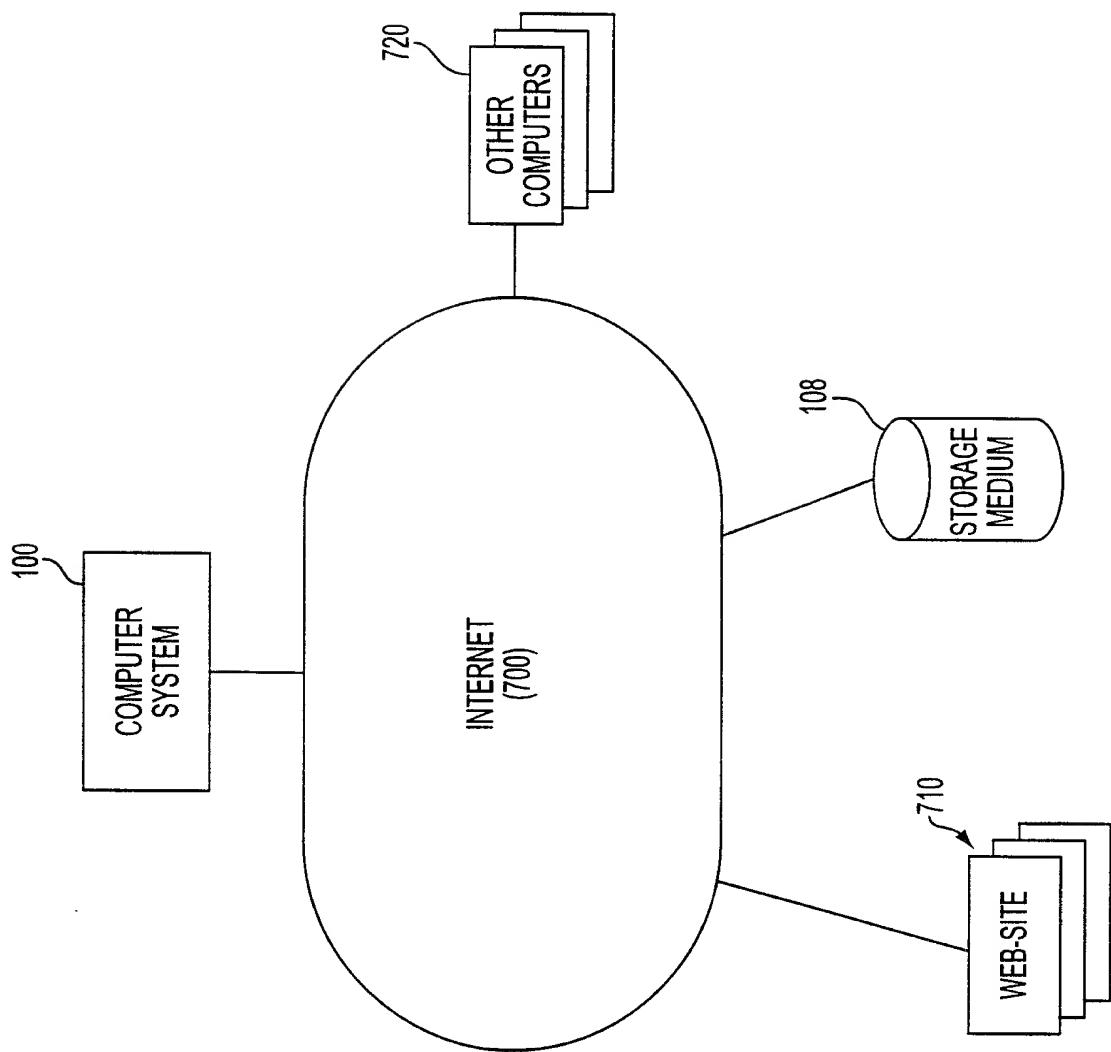


FIG. 8

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/US 99/20447

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 G10L15/26 H04L29/06 G06F17/27 G06F3/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 IPC 7 G10L H04L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 854 418 A (TEXAS INSTRUMENTS INC) 22 July 1998 (1998-07-22) page 2, line 17 - line 23 page 2, line 36 - line 42 page 5, line 52 - line 56 page 6, line 20 - line 25 page 6, line 31 - line 34	1-3, 7-11
Y	—	4-6, 12-14
X	EP 0 747 881 A (AT & T CORP) 11 December 1996 (1996-12-11) column 4, line 12 - line 27 column 5, line 22 - line 54 column 6, line 18 - line 28	1, 7, 8
Y	—	4-6, 12-14
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the International filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the International filing date but later than the priority date claimed

- "T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the International search

15 December 1999

Date of mailing of the International search report

11/01/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax. (+31-70) 340-3016

Authorized officer

Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

Int	ional Application No
PCT/US 99/20447	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 854 417 A (TEXAS INSTRUMENTS INC) 22 July 1998 (1998-07-22) column 3, line 36 -column 4, line 2 column 5, line 1 - line 14 column 5, line 29 - line 37 — DE 44 40 598 C (SIEMENS AG) 23 May 1996 (1996-05-23) abstract column 2, line 46 -column 3, line 17 — WO 98 35491 A (JOHNSTON ROBERT DENIS ;BRITISH TELECOMM (GB)) 13 August 1998 (1998-08-13) column 4, line 28 -column 5, line 8 —	1-14
A		1,2,7-9
A		1,2,7-9

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l Application No

PCT/US 99/20447

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0854418	A	22-07-1998	JP	10240493 A	11-09-1998
EP 0747881	A	11-12-1996	US	5890123 A	30-03-1999
			CA	2177865 A	06-12-1996
			JP	8335160 A	17-12-1996
EP 0854417	A	22-07-1998	JP	10275162 A	13-10-1998
DE 4440598	C	23-05-1996	NONE		
WO 9835491	A	13-08-1998	AU	5674398 A	26-08-1998
			EP	0958692 A	24-11-1999